Automated Modeling of Medical Decisions

John W. Egar and Mark A. Musen
Section on Medical Informatics
Stanford University School of Medicine
Stanford, CA 94305-5479
email: {John_Egar, Mark_Musen}@med.stanford.edu

We have developed a graph grammar and a graph-grammar derivation system that, together, generate decision-theoretic models from unordered lists of medical terms. The medical terms represent considerations in a dilemma that confronts the patient and the health-care provider. Our current grammar ensures that several desirable structural properties are maintained in all derived decision models.

CONSTRUCTION OF MEDICAL DECISION MODELS

In medicine, many important decisions must be made from uncertain evidence. Studies have shown that physicians are subject to erroneous biases when they make decisions based on such information [1]. Decision analysis [2] provides a sound framework—namely, probability theory—for combining uncertain evidence to compute the likelihood of various outcomes. Also, decision analysis provides a basis for weighing explicit patient preferences to derive an optimal plan from the available alternatives. Decision-analytic models, however, can be difficult to compose, and the few models that have been published do not cover the broad spectrum of problems that face physicians every day. Consequently, widespread adoption of normative decision making in medicine depends on the development of improved support for decision modeling.

Previously [3], we have proposed that graph grammars might provide guidance for the automated construction of decision models. In this paper, we report on our implementation of a graph-grammar derivation system, and on our current grammar for medical decision models. The derivation system accepts a list of considerations (tests, diseases, etc.) and generates a model that includes those considerations. We have found that our graph grammar can structure medical decision models that are moderately complex. The grammar can also construct more complex models (i.e., models with over a dozen considerations), but only with help from the user.

QUALITATIVE CONTINGENT INFLUENCE DIAGRAMS

In our research, we represent decision models as **qualitative contingent influence diagrams (QCIDs)**. In this graphical representation, nodes denote variables that are probabilistic (shown as circles), deterministic (double circles), or controlled by the decision maker (squares). A utility node (hexagon) represents the decision maker's value function—a deterministic variable that we wish to maximize by selecting the most propitious decision-node alternatives. Arcs into a decision node delineate information known to the decision maker at the time that the decision is made; these **information** arcs are dashed. Arcs into either probabilistic or deterministic nodes represent probabilistic or functional dependence of the target variable on the source variable; these **dependency** arcs are solid.

Specific probability values are not applicable to all clinical situations, but certain constraints on probability values generalize across cases. For example, curative treatments have negative effects on the probability of diseases. To describe monotonic constraints on probabilistic and functional dependencies, we use **qualitative** arc labels: A "+" restricts the target node's distribution of values to vary in the same direction as changes in the source node's distribution of values, and a "−" restricts the source and target nodes to vary in opposite directions.

GRAPH GRAMMARS

Graph grammars have undergone 20 years of theoretic development [4, 5]. A graph grammar consists of several graph-grammar production rules, which describe syntactic manipulations of a diagram. The particular formalism that we use is

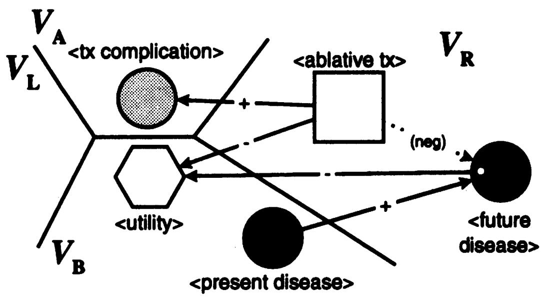

Figure 1: Sample graph-grammar production rule. The production describes how nodes of the type **<ablative tx>** can be added to the host graph. V_L, V_A, V_B, and V_R (left, above, below, and right) are the four regions of a graph-grammar production rule. In this rule, there are no vertices in V_L. (tx = treatment.)

a modification of the operational formalism described by Göttler [6, 7]. In Göttler's formalism, a production can be written as a graph divided into four regions (Figure 1): the left region (V_L), the right region (V_R), the region below (V_B), and the region above (V_A). The two regions V_A and V_B are called the **indeterminate** and **determinate** regions, respectively, and they comprise the **embedding environment** of a rule.

In our system, called **Gramarye,** a production rule describes the following manipulation steps:

1. Locate all regions in the host diagram where the nodes and arcs match the vertices and edges of the determinate and left regions (Figure 2a).*
2. Match zero or more subgraphs in the indeterminate region to subgraphs in the host diagram. Also match the edges between the indeterminate and left regions to corresponding arcs in the diagram.†
3. Delete the nodes that matched V_L, and delete their incident arcs (Figure 2b).
4. Add new nodes and arcs that correspond to the vertices and edges in the right region of the production rule (Figure 2c).

*When V_L and V_B match multiple locations in the host diagram, the derivation system elicits user assistance to choose among the locations.

†When potential matches to V_A exist, the user must determine which nodes should be matched.

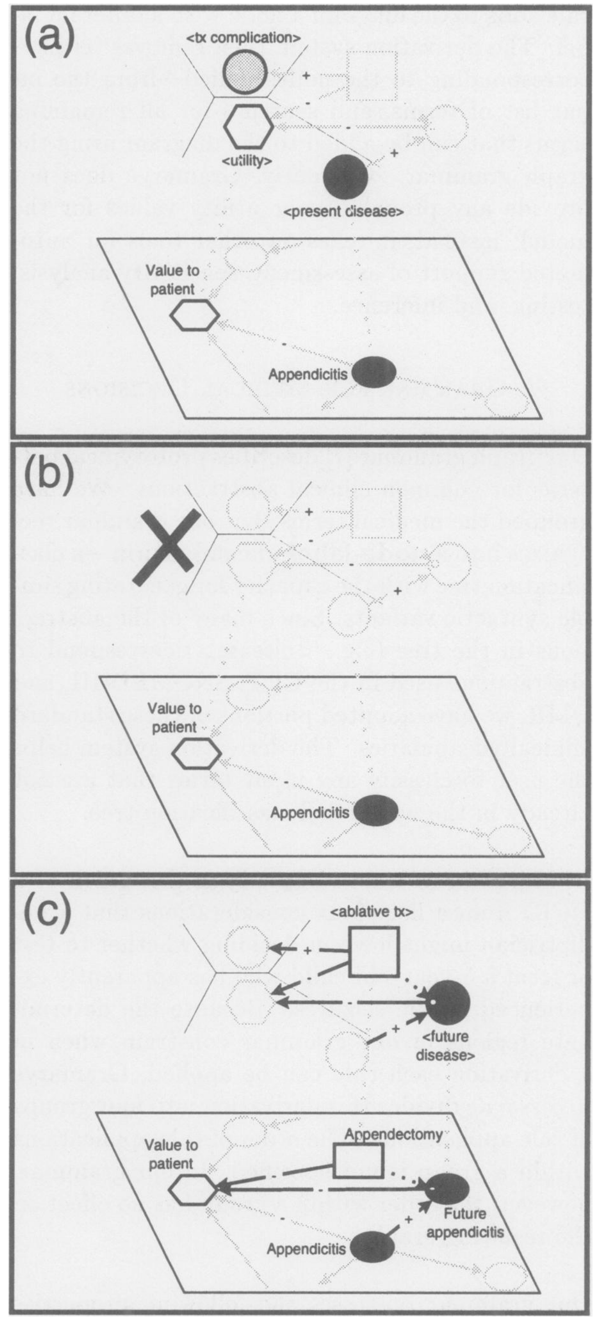

Figure 2: Sample application of the graph-grammar rule from Figure 1. (a) The first view of the host graph shows two nodes from the host diagram matching **<utility>** and **<present disease>** in the production. (b) If V_L contained vertices, a matching set of nodes would be removed. (c) Additional nodes **Appendectomy** and **Future appendicitis** are added to the QCID model.

Gramarye's derivation system invokes all applicable rules as a group, where the application of each rule adds to the diagram a node with a different label. The derivation system then removes terms—corresponding to the nodes added—from the input list of terms, and searches for all remaining terms that can be added to the diagram using the graph grammar. Currently, Gramarye does not provide any probability or utility values for the model; instead, it relies on other tools for automated support of assessment, sensitivity analysis, testing, and inference.

A Grammar for Medical Decisions

Our graph grammar [7] describes prototypical patterns for common clinical abstractions. We have grouped the medical terms that our grammar recognizes into a **node-label classification**—a classification tree with the capacity for generating simple syntactic variants. Since many of the abstractions in the tree (e.g., <**disease**>) correspond to abstractions used in the CPT, SNOMED-III, and QMR, we have adopted portions of these standard clinical vocabularies. The derivation system helps the user to classify any input terms that are not already in the node-label classification tree.

In Figure 3, we show Gramarye's derivation of a QCID from a list of six considerations that a pediatrician might have in deciding whether to test or treat a 5-year-old child who has apparently experienced febrile seizures. Because the determinate regions in our grammar constrain when in a derivation each rule can be applied, Gramarye *necessarily* divides this derivation into four groups of rule applications. The order of rule applications within a group is not specified. In our grammar, however, the order within a group has no effect on the resulting graph.

Our grammar possesses the following properties [8]:

1. The grammar will generate only acyclic graphs.
2. Only one derivation can result from a given input (i.e., the grammar is unambiguous).
3. The grammar will not yield a **qualitatively dominated** decision (i.e., all decisions include both pros and cons).

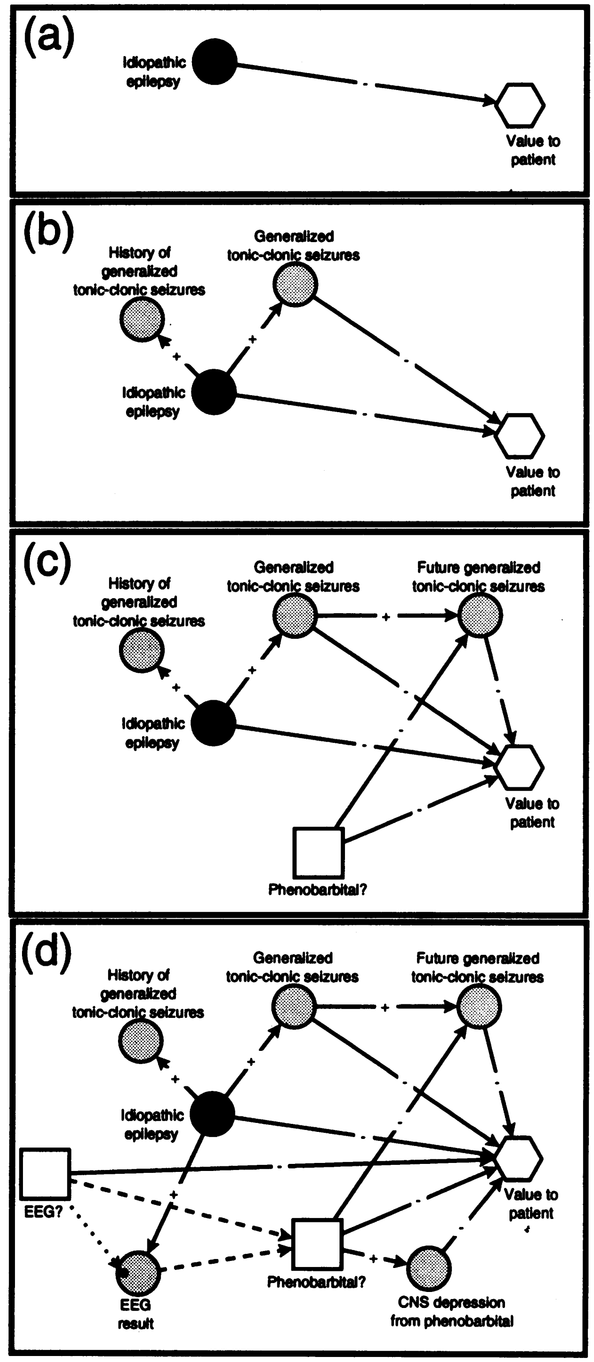

Figure 3: A derivation produced by Gramarye. The terms entered by the user are **generalized tonic-clonic seizures**, **history of generalized tonic-clonic seizures**, **idiopathic epilepsy**, **EEG**, **phenobarbital**, and **CNS depression from phenobarbital**. In part (a), Gramarye triggers a rule from the grammar to add **idiopathic epilepsy**. In (b), Gramarye adds a finding and a malady complication to the model. In (c), the addition of a treatment node entails a future disease node. In (d), Gramarye adds a test and its result. (CNS = central nervous system; EEG = electroencephalography.)

4. Each generated model will contain exactly one utility node.

5. In no generated model will there be successors to the utility node.

6. For all nodes in all generated models, there exists at least one **relevance path** (sequence of solid arcs) to the utility node.

7. For all chance nodes in all generated models, there exists at least one chance-node path to the utility node.

Although these properties do not guarantee that the resulting model is appropriate, they do prevent common modeling errors. For example, the third property—no dominated decisions—accounts for five of the nine rules used by Wellman and colleagues to critique manually composed decision trees [9]. The other four critiquing rules in Wellman's system are obviated by our adoption of the influence diagram notation.

RESULTS

We have implemented Gramarye in Common Lisp. We have also implemented a graphical user interface in NeXTSTEP, wherein the user selects or enters terms corresponding to considerations for a given decision problem, and then views and may edit the QCID that Gramarye generates automatically. Our grammar—both the node-label classification and the production rules—has evolved continuously over the past 6 months. Our derivation system has remained fairly stable. When presented with training cases, Gramarye has produced models that are roughly equivalent (i.e., mathematically identical, once a single utility function is assessed) to those developed by independent researchers. For example, other researchers [10] have modeled a complex clinical decision regarding coronary-artery bypass graft, abdominal aortic aneurysm, cardiac catheterization, and 14 other considerations for an elderly patient. By entering the same list of considerations into Gramarye, we can derive an equivalent qualitative model in roughly 1 minute. Professional decision analysts have found that the initial qualitative modeling of a decision problem often requires more effort than does the assessment of probabilities and utilities for the model. Consequently, our approach to modeling could make normative decision making significantly more accessible to health-care workers.

We have not yet had sufficient experience with our system to warrant a more formal evaluation with test cases. However, we have discovered five important problems with our approach:

1. Our vocabulary of roughly 6000 terms is small when compared to the variety of words used to denote considerations involved in general internal medicine. However, nonmedical considerations often play an important role in clinical decisions. Users of our system will need support in classifying these considerations according to abstractions in the node-label tree.

2. Although most terms that we have considered fall under a single classification in our node-label tree, some terms have more than one role, depending on the context of the decision problem. For example, an excisional biopsy may be considered as both a `<test>` and as a potential `<treatment>` for a small skin lesion. Currently, when an entered term appears in multiple places in the classification tree, the derivation system requires the user to choose among the possible roles for that term. Concepts—such as excisional biopsy—that present features of more than one abstraction in a single decision problem may require new abstractions and new productions in our grammar.

3. As exceptions to our existing patterns arise, the grammar continues to evolve. Establishing that the seven properties discussed in this paper hold for each new grammar can be tedious. Because the properties are neither obvious nor easy to prove, we are working to devise a framework—much like the parse tables used in context-free grammars—that would enable us to evaluate automatically a grammar for such properties.

4. As we relax assumptions that are inherent in the current grammar (e.g., that each treatment is administered for a single disease), Gramarye's requirement for user assistance grows, and the guidance provided by static models of the domain becomes increasingly desirable. By introducing purely semantic distinctions (e.g., dividing respiratory findings, diseases, and tests from other findings, diseases, and tests), we can reduce the nondeterminism that results when Gramarye can

add the same node at multiple locations. Unfortunately, most findings and diseases are difficult to partition into neat categories.

5. The current grammar cannot construct intricate chance-node subgraphs that reflect physiologic understanding. Because of this shortcoming, we have begun to use static belief-network models of human pathophysiology to guide the derivation process. We have not yet had sufficient experience with this knowledge-base supplementation of the grammar to estimate its benefit. The only alternative that we see to providing static models of physiologic relationships is imbedding such medical knowledge into the grammar, and we suspect that such a grammar would become unmanageable. Chance-node abstractions for physiologic parameters will probably emerge as our grammar evolves, but we have yet to see how such abstractions will augment the grammar's modeling power.

Additional graph-grammar productions may help the user to decide how to simplify the value function when the generated model contains numerous arcs into the utility node. Assumptions of causal independence would greatly simplify probability assessment for nodes with many parents, but we have not yet found general patterns of causal independence that we could embed in the grammar.

Discussion

By generating QCIDs from lists of considerations, Gramarye has shown that graph grammars can help physicians to model decision problems. Such models provide a basis for balancing patient preferences, individual features of a clinical case, and statistical data from the literature to arrive at a sound, responsive, and cost-effective decision. The strength of our approach lies in our ability to change a constructive task—where the user must decide to include or omit in her model each of an exponential number of possible relationships—to a classification task, which grows linearly with the number of considerations (terms) to be included in the model.

Acknowledgments

We are grateful to Jack Breese, Keith Campbell, Henrik Eriksson, David Heckerman, Ross Shachter, Edward Shortliffe, and Jaap Suermondt for their support and guidance.

We conducted this work with the support of the National Library of Medicine under grants LM-05157 and LM-07033. Computing support was provided by the CAMIS resource, funded under grant number LM-05305 from the National Library of Medicine. Dr. Musen is the recipient of a National Science Foundation Young Investigator Award.

Reference

[1] A. Tversky and D. Kahneman. Judgement under uncertainty: Heuristics and biases. *Science*, 185:1124–1131, 1974.

[2] H. Raiffa. *Decision Analysis.* Addison-Wesley, Reading, MA, 1968.

[3] J. Egar and M. Musen. Graph-grammar productions for modeling medical dilemmas. In M. Frisse, editor, *Proceedings of the 16th SCAMC*, pages 349–353, Baltimore, MD, November 1992. McGraw-Hill, Inc.

[4] J. Feder. Plex languages. *Information Science*, 3:225–241, 1971.

[5] H. Ehrig, H-J. Kreowski, and G. Rozenberg, editors. *Graph Grammars and Their Application to Computer Science: Proceedings of the Fourth International Workshop*, Berlin, 1991. Springer-Verlag.

[6] H. Göttler. Diagram editors = graphs + attributes + graph grammars. *International Journal of Man–Machine Studies*, 37:481–502, 1992.

[7] John W. Egar and Mark A. Musen. Graph-grammar assistance for automated generation of influence diagrams. In David Heckerman and Abe Mamdani, editors, *Uncertainty in Artificial Intelligence: Proceedings of the Ninth Conference*, pages 235–242, Washington, D.C., July 1993. Morgan Kaufmann.

[8] J. Egar. Properties maintained by a graph grammar for medical decisions. Technical Report KSL-93-25, Knowledge Systems Laboratory, Stanford University, Stanford, CA, March 1993.

[9] M. Wellman, M. Eckman, C. Fleming, S. Marshall, F. Sonnenberg, and S. Pauker. Automated critiquing of medical decision trees. *Medical Decision Making*, 9:272–284, 1989.

[10] Michael P. Wellman. *Formulation of Tradeoffs in Planning Under Uncertainty.* Morgan Kaufmann, San Mateo, CA, 1990.